

# Morphological Lightcurve Distortions due to Finite Integration Time

David M. Kipping<sup>1,2\*</sup>

<sup>1</sup>*Harvard-Smithsonian Center for Astrophysics,  
60, Garden Street, Cambridge, MA 02144, USA*

<sup>2</sup>*Department of Physics and Astronomy, University College London,  
Gower Street, London WC1E 6BT, UK*

Accepted . Received ; in original form

## ABSTRACT

We explore how finite integration times or equivalently temporal binning induces morphological distortions to the transit lightcurve. These distortions, if uncorrected for, lead to the retrieval of erroneous system parameters and may even lead to some planetary candidates being rejected as ostensibly unphysical. We provide analytic expressions for estimating the disturbance to the various lightcurve parameters as a function of the integration time. These effects are particularly crucial in light of the long-cadence photometry often used for discovering new exoplanets by for example *CoRoT* and the *Kepler Mission* (8.5 and 30 minutes). One of the dominant effects of long integration times is a systematic underestimation of the lightcurve derived stellar density, which has significant ramifications for transit surveys. We present a discussion of numerical integration techniques to compensate for the effects and produce expressions to quickly estimate the errors of such techniques, as a function of integration time and numerical resolution. This allows for an economic choice of resolution before attempting fits of long-cadence lightcurves.

**Key words:** techniques: photometric — planets and satellites: general — planetary systems — occultations

## 1 INTRODUCTION

Transiting extrasolar planets have become a powerful tool since the first discovery by Charbonneau et al. (2000) and Henry et al. (2000). The detection of a transit allows for a measurement of the planet-star ratio-of-radii, which may be used to infer the planetary radius once the stellar radius has been determined. Multi-wavelength measurements may be used to retrieve a spectrum of the light being absorbed by the planetary atmosphere and thus a deduction of the molecular constituents (e.g. Tinetti et al. 2007). Timing of the eclipse minima and duration permits for the detection of other bodies in the system (Agol et al. 2005; Holman et al. 2005) and companion moons (Kipping 2009a,b). In recent years, various authors have published increasingly ingenious methods of characterizing exoplanets through transits, including thermal emission detections (Deming et al. 2006), spin-orbit alignment (Winn et al. 2005), oblateness measurements (Seager & Hui 2002; Carter et al. 2010) and many more.

As we look for these small whispers of the exoplanet's nature, we are in fact examining the morphology of the transit lightcurve in increasing detail. It is therefore paramount that the methods for modeling these lightcurves are in the most precise form possible.

In this paper, we explore the effects of finite integration time on the transit lightcurve, or equivalently the act of temporal binning. Long-cadence (LC) data smears out the transit lightcurve signal into a broader shape which will lead to an erroneous retrieval of the system parameters, if unaccounted for.

We will first outline the nature of the morphological distortions in §2 and then derive analytic expressions for the approximate systematic error in the retrieved parameters, as a function of integration time. It is shown that the effect can be significant enough to even lead to some planetary candidates being rejected as being unphysical, unless the effect is correctly accounted for. Obviously, this is critical for missions like *CoRoT* and *Kepler* which employ LC photometry to detect new transiting planets. In §3, we will discuss numerical techniques to correctly model the transit lightcurve and derive the errors of these methods as a function of in-

\* E-mail: dkipping@cfa.harvard.edu

tegration time and numerical resolution. These expressions allow for a simple evaluation of the expected consequences of finite integration time and thus will guide observers in choosing how to deal with this issue.

## 2 THE EFFECTS OF FINITE INTEGRATION TIME

### 2.1 Ingress/Egress Durations

For a transit lightcurve, there are four critical contact points which define the overall shape, which represent the points where the time derivative is discontinuous. Physically speaking, contact points I and IV occur when the sky-projected planet-star separation is equal to the stellar radius plus to planetary radius. Contact points II and III occur when this parameter equals the stellar radius minus the planetary radius. Defining  $W$  as the average of the durations between the 1<sup>st</sup>-to-4<sup>th</sup> and 2<sup>nd</sup>-to-3<sup>rd</sup> contacts (Kipping et al. 2010),  $t_c$  as the mid-transit time and  $\tau$  as the ingress/egress duration, we have:

$$t_I = t_c - W/2 - \tau/2 \quad (1)$$

$$t_{II} = t_c - W/2 + \tau/2 \quad (2)$$

$$t_{III} = t_c + W/2 - \tau/2 \quad (3)$$

$$t_{IV} = t_c + W/2 + \tau/2 \quad (4)$$

The principal effect of finite integration time is to smear out the lightcurve into a broader shape (see figure 1). The apparent ingress and egress durations will increase and additional curvature will be introduced into the lightcurve wings. The ingress/egress stretching can be considered in terms of the apparent positions of the contact points being temporally shifted from their true value. The magnitude of this time shift is dependent on the relative phase difference between the sampling and the transit signal. If we assume that a large number of transits observed with LC photometry are folded about the orbital period, as is typical in transit detection, then the effect becomes much more predictable with the deviation averaging out to  $\mathcal{I}/2$ .

Under these conditions, contact points I and IV will appear to move outwards from  $t_c$  by one half of the integration time each,  $\mathcal{I}/2$ . Conversely, contact points II and III will appear to move inwards by the same amount. Let us define the apparent contact points as  $t'_I$ :

$$t'_I = t_I - \mathcal{I}/2 \quad (5)$$

$$t'_{II} = t_{II} + \mathcal{I}/2 \quad (6)$$

$$t'_{III} = t_{III} - \mathcal{I}/2 \quad (7)$$

$$t'_{IV} = t_{IV} + \mathcal{I}/2 \quad (8)$$

Seager & Mallén-Ornelas (2003) showed that several physical system parameters may be retrieved by analysis of the lightcurve morphology. Specifically, they define  $t_T = t_{IV} - t_I$  and  $t_F = t_{III} - t_{II}$ . Let us define the apparent observed values as  $t'_T$  and  $t'_F$  respectively. Seager & Mallén-Ornelas (2003) showed that the semi-major axis of the planetary orbit divided by the stellar radius,  $a_R$ , the transit impact parameter,  $b$ , and the ratio-of-radii,  $p$ , are all determined by manipulation of  $t_T$ ,  $t_F$  and

transit depth,  $\delta$ . These parameters may then be used to determine orbital inclination,  $i$ , and stellar density,  $\rho_*$ . Recently, Kipping (2010) extended these equations to account for orbital eccentricity and we will here employ these more general expressions in our analysis:

$$b^2 = \frac{(1-p)^2 - \frac{\sin^2[(t_F \pi \sqrt{1-e^2})/(P \varrho_c^2)]}{\sin^2[(t_T \pi \sqrt{1-e^2})/(P \varrho_c^2)]} (1+p)^2}{1 - \frac{\sin^2[(t_F \pi \sqrt{1-e^2})/(P \varrho_c^2)]}{\sin^2[(t_T \pi \sqrt{1-e^2})/(P \varrho_c^2)]}} \quad (9)$$

$$a_R^2 = \frac{(1+p)^2 - b^2}{\varrho_c^2 \sin^2[(t_T \pi \sqrt{1-e^2})/(P \varrho_c^2)]} + \frac{b^2}{\varrho_c^2} \quad (10)$$

$$\rho_* = \frac{3\pi}{GP^2} a_R^3 - p^3 \rho_P \quad (11)$$

We here derive what these retrieved parameters would be if the effect of finite integration time was ignored. In the above expressions, we replace  $t_T$  &  $t_F$  with  $t'_T = (t_T + \mathcal{I})$  &  $t'_F = (t_F - \mathcal{I})$ . The new expressions then have the remaining  $t_T$  &  $t_F$  terms written out in terms of the true values of  $b$ ,  $a_R$ , etc. This leaves us with  $b'(b, a_R, \dots)$ ,  $a'_R(b, a_R, \dots)$ , etc:

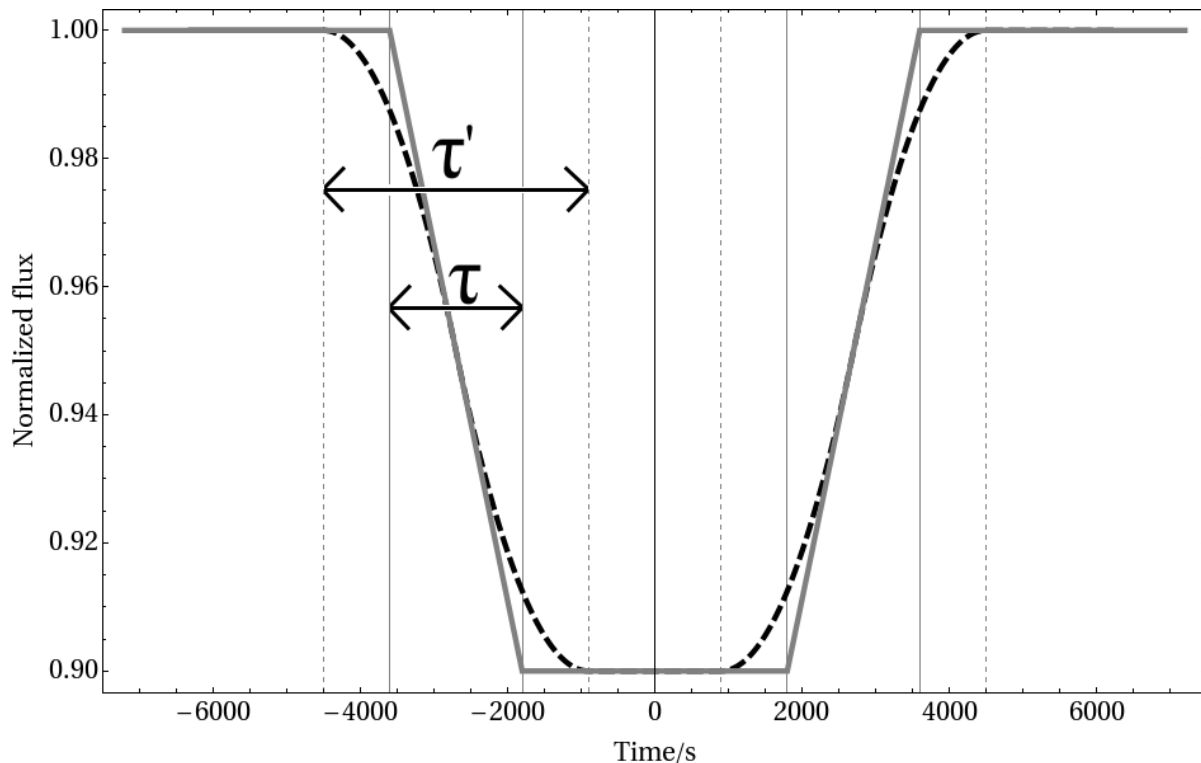
$$b'^2 = \frac{\left[ (1+p)^2 \frac{\sin^2 \left[ \frac{\sqrt{1-e^2} \mathcal{I} \pi - \arcsin \left( \frac{\sqrt{(1-p)^2 - b^2}}{a_R \varrho_c \sin i} \right)}{P \varrho_c^2} \right]}{\sin^2 \left[ \frac{\sqrt{1-e^2} \mathcal{I} \pi + \arcsin \left( \frac{\sqrt{(1+p)^2 - b^2}}{a_R \varrho_c \sin i} \right)}{P \varrho_c^2} \right]} - (1-p)^2 \right]}{\left[ \frac{\sin^2 \left[ \frac{\sqrt{1-e^2} \mathcal{I} \pi - \arcsin \left( \frac{\sqrt{(1-p)^2 - b^2}}{a_R \varrho_c \sin i} \right)}{P \varrho_c^2} \right]}{\sin^2 \left[ \frac{\sqrt{1-e^2} \mathcal{I} \pi + \arcsin \left( \frac{\sqrt{(1+p)^2 - b^2}}{a_R \varrho_c \sin i} \right)}{P \varrho_c^2} \right]} - 1 \right]} \quad (12)$$

$$a'^2_R = \frac{\left[ (1+p)^2 \sin^2 \left( \frac{\sqrt{1-e^2} \pi \mathcal{I}}{\varrho_c^2 P} - \arcsin \left( \frac{\sqrt{(1-p)^2 - b^2}}{a_R \varrho_c \sin i} \right) \right) - 4p \right]}{\sin^2 \left( \frac{\sqrt{1-e^2} \pi \mathcal{I}}{\varrho_c^2 P} + \arcsin \left( \frac{\sqrt{(1+p)^2 - b^2}}{a_R \varrho_c \sin i} \right) \right)} - \frac{\sin^2 \left( \frac{\sqrt{1-e^2} \pi \mathcal{I}}{\varrho_c^2 P} - \arcsin \left( \frac{\sqrt{(1-p)^2 - b^2}}{a_R \varrho_c \sin i} \right) \right)}{\varrho_c^2 \sin^2 \left( \frac{\sqrt{1-e^2} \pi \mathcal{I}}{\varrho_c^2 P} + \arcsin \left( \frac{\sqrt{(1+p)^2 - b^2}}{a_R \varrho_c \sin i} \right) \right)} - \varrho_c^2 \quad (13)$$

$$\rho'_* \simeq \frac{3\pi}{GP^2} a'^3_R \quad (14)$$

Setting  $\mathcal{I} = 0$  returns the original results as expected. Unfortunately, these equations are somewhat overly complex for us to draw any physical intuition. To proceed, let's consider a typical case example by using the system parameters from one of the *Kepler* planets, since these are discovered using long-cadence data. The following example is for the assumption of zero limb darkening, which is a very poor one for the *Kepler* bandpass. The effects of limb darkening will be discussed later.

In figure 2, we plot the retrieved stellar density,  $\rho'_*$ , as a function of the true stellar density,  $\rho_*$ . We fix all other parameters to be that of *Kepler*-5b, as reported by Koch et al. (2010). The effect can be seen to be highly significant, causing the retrieved stellar density to be underestimated by a factor which borders on an order-of-magnitude. This scale of underestimation is sufficient to completely reject some plan-



**Figure 1.** A trapezoid approximated lightcurve with a one hour flat-bottom duration and 30 minute ingress/egress duration,  $\tau$ , is shown in solid. The dashed line shows the lightcurve morphology for an integration time of 30 minutes. The apparent ingress/egress duration,  $\tau'$ , can be seen to have doubled purely as a consequence of the integration time.

etary candidates as unphysical. However, we note that in reality the underestimation of  $\rho_*$  will not be this severe due to countering effects of limb darkening suppression discussed later in §2.2.

Looking at figure 2 again, let us explore the physical reasons for this underestimation. We begin our line of thought by considering the impact parameter,  $b$ . If the ingress/egress duration is elongated, what do we expect to happen to the derived impact parameter? Consider two extreme cases. When a planet transits with a very low impact parameter, we have essentially an equatorial transit. This means the vector describing the sky-projected planetary velocity is nearly perpendicular to the stellar limb. As a result, the planet crosses the limb expediently. In contrast, if we have a near-grazing transit, the mutual angle between the stellar limb and the sky-projected planetary velocity vector has become more acute, which has the effect of making the limb-crossing time much longer. Therefore, stretching the ingress/egress duration causes  $b' > b$ .

Having established this point, consider the effect on  $a_R$ . The simplest way to understand the effect on this parameter is to appreciate that  $b$  and  $a_R$  exhibit an extremely strong negative correlation, as demonstrated by Carter et al. (2008). So the act of increasing the ingress/egress duration will increase  $b$  and therefore decrease  $a_R$ . Finally, from equation (11) we know that  $\rho_* \propto a_R^3$ .

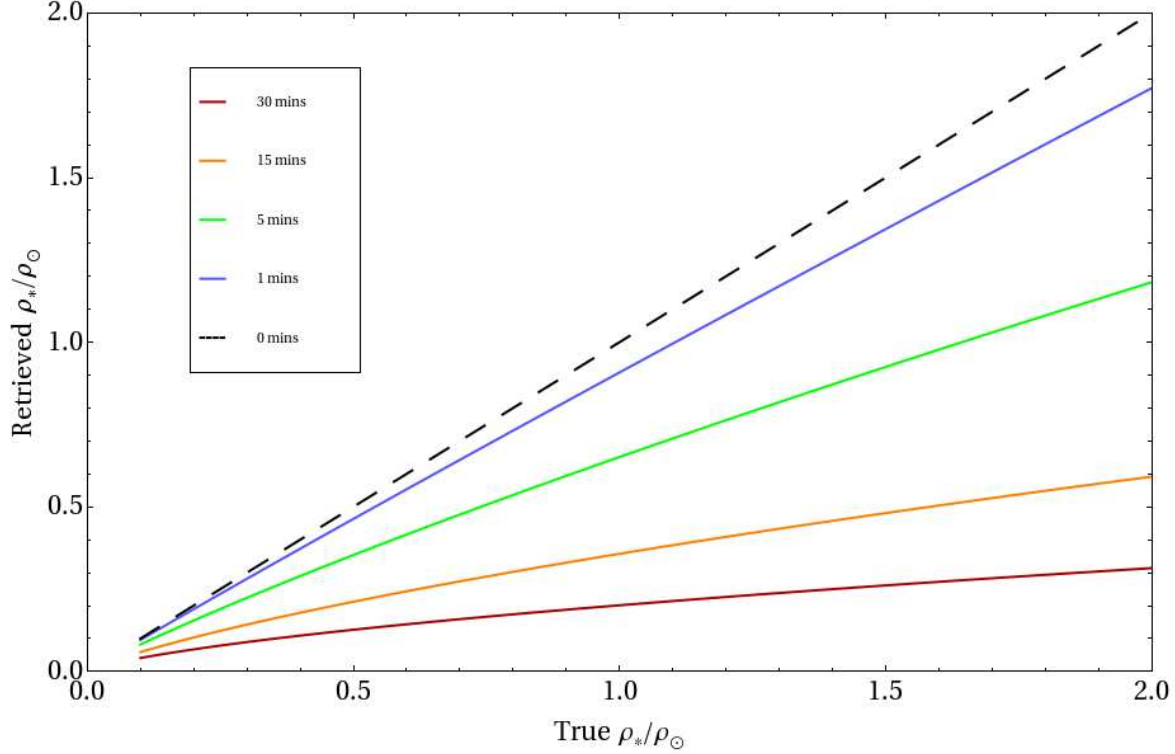
## 2.2 Limb Darkening Effects

For transit observations at visible wavelengths, limb darkening is quite pronounced producing a well-known curvature in the flat-bottom part of the lightcurve. For large integration times, the curvature is smeared out, producing a flatter transit trough morphology.

If we were trying to fit such a transit lightcurve, and the limb darkening parameters are exactly known, our hypothetical fitting algorithm would move towards a case where limb-darkening is least effective. This occurs for near-equatorial transits. This is because as the transit impact parameter becomes larger and larger, we move closer and closer to the limb of the star where limb darkening accentuates. Therefore, LC data pushes  $b'$  towards a more equatorial value due to limb darkening i.e.  $b' < b$ .

## 2.3 Two Countering Effects

In conclusion, finite integration times have two countering effects on  $b'$  and thus  $a'_R$  and  $\rho'_*$  as well. The ingress and egress smearing causes the ingress to appear larger, which occurs for more grazing transits. In contrast, the limb darkening is suppressed, which occurs for more equatorial transits. Whilst limb darkening is important, especially for transit surveys like *CoRoT* and *Kepler* which operate at visible wavelengths, the fundamental change in the transit morphology is sufficiently large that it will tend to dominate. This is because the amplitude of the curvature in the lightcurve trough is usually at least an order-of-magnitude less than



**Figure 2.** As an example, we use the system parameters of *Kepler-5b* to show the effect on the retrieved stellar density as a result of long integration times, in the case of no limb darkening. The 1 min line appears to produce results within the typical uncertainties of the derived stellar density.

the amplitude of the transit signal itself. Therefore, the consequences of smearing the overall transit signal will tend to dominate over the suppression of limb darkening.

The general rule of thumb that  $b'$  is overestimated will break-down when we have  $\tau \gg \mathcal{I}$  but  $t_F \sim \mathcal{O}[\mathcal{I}]$ , which occurs for near-grazing transits. In such a case, the fractional change to the ingress durations is minimized but the change in the transit trough curvature is maximized.

To exactly calculate the net consequence of these two effects though, the integration time should be included when we generate our model lightcurves rather than attempting ad-hoc corrections post-analysis. This appears to be the only way to completely account for the effect in a reliable manner.

#### 2.4 Consequences for the Transit Depth

We will briefly comment on the effect of finite integration time on the transit depth. Assuming no limb darkening is present and  $t_F > \mathcal{I}$ , the transit depth is completely unaffected by the long integration time. This has important consequences for secondary eclipses where the lightcurve is unaffected by stellar limb darkening.

For cases where we have limb darkening, the net effect on the retrieved  $p'$  will depend on whether  $b' > b$  or  $b < b'$ . Additionally, the effect will be a function of what assumptions were used in the fitting algorithm (e.g. the fixing of various parameters), the true impact parameter and the limb darkening coefficients. Given the large number of correlated factors, predicting the effect of integration time becomes less reliable and we must fit the transit lightcurve

with a model which accounts for integration time in the first place.

#### 2.5 Observed Effects with *Kepler*

We point out that the effects of long integration times have already been observed by the *Kepler Mission*. Figure 4 of Gilliland et al. (2010), shows the transit lightcurve of long-cadence and short-cadence data for the same planet, TrES-2b. The long-cadence lightcurve exhibits a broader shape with the apparent position of the contact points shifted by  $\sim \mathcal{I}/2$ , as predicted by our model. Notice also that the curvature in the transit trough, due to limb darkening, has also been attenuated.

### 3 ACCURATE TRANSIT LIGHTCURVE MODELING

#### 3.1 Analytic Integration

The critical problem we have outlined can be simply summarized by the following: *Don't fit an unbinned model to binned data*. The model usually used to generate a transit lightcurve is provided by Mandel & Agol (2002) (MA02), which includes the effects of stellar limb darkening. To generate the MA02 lightcurve, we usually have a set of time stamps forming a time vector  $\underline{t}$ . This time vector represents instantaneous moments rather than integrated time. The  $\underline{t}$  vector is converted into a vector of instantaneous true

anomalies,  $\underline{f}$ , by solving Kepler's equation numerically.  $\underline{f}$  is converted to a  $\underline{z}$  array, where  $z(f)$  is the instantaneous sky-projected planet-star separation. Finally, the MA02 equations provide us with  $\underline{F}$ , where  $F$  is the instantaneous flux. We summarize the sequence of events as:

$$\underline{t} \rightarrow \underline{f} \rightarrow \underline{z} \rightarrow \underline{F}(\underline{t})$$

Now that we have established the mechanism of generating a transit lightcurve for instantaneous time stamps,  $\underline{F}(\underline{t})$ , let us consider what the transit lightcurve for integrated time stamps would be,  $\tilde{F}(\tilde{t})$ . In this case, the integrated flux would be given by:

$$\tilde{F}(\tilde{t}) = \frac{\int_{t=\tilde{t}-\mathcal{I}/2}^{\tilde{t}+\mathcal{I}/2} F(t) dt}{\int_{t=\tilde{t}-\mathcal{I}/2}^{\tilde{t}+\mathcal{I}/2} dt} \quad (15)$$

This equation suffers from the problem that  $\underline{F}$  cannot be written as a function of  $t$  analytically, since such a solution would require a closed-form solution to Kepler's equation, which is transcendental. Evaluating this expression for  $\underline{F}$  as a function of  $z$  is also not possible since we would find the following:

$$\int_{t=\tilde{t}-\mathcal{I}/2}^{\tilde{t}+\mathcal{I}/2} F(t) dt = \int_{z(\tilde{t}-\mathcal{I}/2)}^{z(\tilde{t}+\mathcal{I}/2)} F(t(z)) \frac{dt}{dz}(z) dz \quad (16)$$

Whilst  $dz/dt$  may be evaluated analytically through a chain rule expansion of  $(dz/df) \times (df/dt)$ , the resultant expression will be as a function of  $f$ , rather than  $z$ . If we knew  $f(z)$ , then we would be able to write out the integrand in a closed-form, but  $f(z)$  can only be found by solving a quartic equation, as shown by Kipping (2008) (K08). Unfortunately, as discussed in Kipping (2010), there is no currently proposed method to correspond which roots refer to which orbital conjunction which makes a closed-form expression elusive.

The only remaining hope for a simple analytic expression would be to express the integral in terms of true or eccentric anomaly, which are inter-changeable. This would yield the following integral:

$$\int_{t=\tilde{t}-\mathcal{I}/2}^{\tilde{t}+\mathcal{I}/2} F(t) dt = \int_{f(\tilde{t}-\mathcal{I}/2)}^{f(\tilde{t}+\mathcal{I}/2)} F(t(f)) \frac{dt}{df}(f) df \quad (17)$$

The integrand of this expression may be written out in a closed-form, by utilizing the solutions of K08:

$$\int_{t=\tilde{t}-\mathcal{I}/2}^{\tilde{t}+\mathcal{I}/2} F(t) dt = \int_{f(\tilde{t}-\mathcal{I}/2)}^{f(\tilde{t}+\mathcal{I}/2)} F(z(f)) \frac{P}{2\pi} \frac{1}{\sqrt{1-e^2}} D(f) df \quad (18)$$

Where  $D(f)$  is the duration function defined by K08,  $F(z)$  is given by MA02 and  $z(f)$  is well-known (e.g. Winn (2010)). The integral limits do not possess a closed-form solution since once again we must solve Kepler's equation, but in principle the indefinite integral could be analytically evaluated and then the relevant limits applied after a sub-routine provides numerical solutions to Kepler's equation. We believe that this strategy would be the most computationally efficient since we have obviated the need for any

numerical integration. However, we were unable to find a solution for the indefinite integral for even a uniform-source case and will therefore focus the remainder of this discussion onto numerical techniques.

### 3.2 Numerical Integration

Having established the significant challenges regarding analytic integration, we now turn our attention to the use of numerical integration techniques. The functions we need to integrate over are in fact very well-behaved and well-approximated by compositions of polynomials and thus we anticipate even a low-resolution numerical integration technique should provide satisfactory accuracy.

In this subsection, we first consider the merit of Simpson's Rule or other Newton-Cotes based methods. We aim to avoid using nested quadrature methods like the Gauss-Kronrod or Clenshaw-Curtis, as the number of integrations required is large and we wish to avoid nested methods. For the simplest case of Simpson's rule, we have:

$$\tilde{F}_i(N=3) = \frac{F(t_i - \mathcal{I}/2) + 4F(t_i) + F(t_i + \mathcal{I}/2)}{6} \quad (19)$$

Where  $N$  denotes the number of calls needed to the MA02 code and essentially is a measure of the resolution of our numerical integration. This method may be extended to higher intervals by using Simpson's composite rule. Alternatively, we can extend to cubic, quartic, etc interpolations by using the Newton-Cotes formulas. Each time we evaluate  $F(t)$  requires another call to the MA02 subroutine, and thus we wish to minimize the number of calls, but maximize the accuracy of the employed technique.

Simpson's composite rule works by splitting up our integration range into  $2m$  subintervals, therefore requiring  $N = 2m + 1$  calls to the MA02 code. The error on the composite Simpson's rule is given by:

$$\sigma_{\tilde{F}}(N)_{\text{Comp.Simpson}} = F^{(4)}(\epsilon) \frac{\mathcal{I}}{180} \left( \frac{\mathcal{I}}{N-1} \right)^4 \quad (20)$$

Where  $\epsilon$  is some number between  $t_i - \mathcal{I}/2$  and  $t_i + \mathcal{I}/2$ . In contrast, the Newton-Cotes formulas move through increasing orders by increasing the interpolation order. For the  $N = 4$  case (which is the cubic interpolation scenario, known as Simpson's 3/8 rule), the equivalent errors between the two methods are:

$$\sigma_{\tilde{F}}(N=4)_{\text{Newton-Cotes}} = F^{(4)}(\epsilon) \frac{3}{80} \mathcal{I}^5 \quad (21)$$

$$\sigma_{\tilde{F}}(N=4)_{\text{Comp.Simpson}} = F^{(4)}(\epsilon) \frac{1}{2880} \mathcal{I}^5 \quad (22)$$

Thus for  $N = 4$ , Simpson's composite rule offers greater accuracy than the Newton-Cotes based equation. Moving through the higher orders in the Newton-Cotes family causes the error to have a functional dependence on  $F^{(N)}$ , i.e. the  $N^{\text{th}}$  differential of  $F$ . So for  $N > 4$  it is not possible to give an exact comparison between the two methods since  $F^{(N)}(t)$  is not known for any  $N$ . Therefore, our only reliable comparison is for the  $N = 4$  case, from which we conclude the composite Simpson's rule is superior in terms of accuracy versus computational requirement.

### 3.3 Error in Numerical Integration

Let us now consider what value of  $N$  should be used. There are essentially three segments of the lightcurve which exhibit curvature and thus would produce the maximum error in our numerical integrations, which employ linear piece-wise approximations.

- (i) Curvature of the ingress/egress
- (ii) Curvature of the limb-darkened lightcurve trough
- (iii) Discontinuities at the contact points

The last of these is due to a discontinuous function and the former two are due to curvatures within continuous functions. We will treat these two different sources of ‘curvature’ separately, although from the arguments made earlier, we expect the last of these effects to be the largest source of numerical error.

#### 3.3.1 Ingress/egress curvature

The transit lightcurve has a depth  $\delta$  and an ingress duration  $\tau$ . For most of the ingress, the curvature is close to zero and essentially mimics a linear slope. However, near the contact points, the slope rapidly changes to a flat line of zero gradient. Therefore, near the contact points, the ingress/egress morphology causes large amounts of curvature. These points will exhibit the largest numerical errors in using a technique like Simpson’s composite rule.

A suitable choice of resolution can be made by increasing  $N$  until  $\sigma_{\hat{F}}|_{\max} \leq \sigma_{\hat{F},\text{obs}}$ , i.e. our calculation should produce a flux which has a maximum systematic error which is less than the observational uncertainty. We will set our resolution to a point where it provides satisfactory accuracy even at the point of highest numerical error, i.e. within the ingress/egress near the contact points.

Another approach would be to use an adaptive composite Simpson’s rule, for example like that proposed by McKee (1962). However, our preference here is to avoid using adaptive routines since they would require a new adaption for every single data point and fitting trial, which would be time consuming. The costs versus benefits of using such a method could warrant further investigation in the future. Instead, we choose to use the adaption required for the most troublesome points, which we have already identified. The required interval size in each element of the Simpson’s composition should be decreased until we reach:

$$|S(a, \frac{a+b}{2}) + S(\frac{a+b}{2}, b) - S(a, b)|/15 < \sigma_{\hat{F},\text{obs}} \quad (23)$$

Where  $S(\alpha, \beta)$  is Simpson’s rule evaluated over the interval  $\alpha$  to  $\beta$ . In our case, the integral is over time and  $a = t_I$  and  $b = t_I + (\mathcal{I}_0/m)$ , where  $2m$  is the number of subintervals we split the integral into and  $2m = N - 1$  where  $N$  is the required factor by which the number of calls to the MA02 code increases by. The reason for the subscript of 0 by the  $\mathcal{I}$  term will be explained shortly. Our requirement may be written as:

$$\frac{|S(t_I, t_I + \frac{\mathcal{I}_0}{2m}) + S(t_I + \frac{\mathcal{I}_0}{2m}, t_I + \frac{\mathcal{I}_0}{m}) - S(t_I, t_I + \frac{\mathcal{I}_0}{m})|}{15} < \sigma_{\hat{F},\text{obs}} \quad (24)$$

$$S(\alpha, \beta) = \frac{\beta - \alpha}{6} \left[ F(\alpha) + 4F\left(\frac{\alpha + \beta}{2}\right) + F(\beta) \right] \quad (25)$$

In order to continue, we need to evaluate  $F(t)$  in a closed-form, which cannot be achieved due to Kepler’s equation. However, there exists a special case where Kepler’s equation does yield an exact closed-form solution and this occurs for circular orbits since  $M = E = f$ . In such a case, we may write:

$$z(t) = a_R \sqrt{\sin^2\left(\frac{2\pi t}{P}\right) + \cos^2 i \cos^2\left(\frac{2\pi t}{P}\right)} \quad (26)$$

The ingress/egress morphology is dominated by the expressions pertaining to a uniform source. Limb darkening does affect the ingress/egress curvature but this is much less than the amplitude of the uniform source transit signal. In the small-planet limit, MA02 provided the following approximation for the ingress/egress flux:

$$F(x) = 1 + x\sqrt{p^2 - x^2} - p^2 \arccos\left[\frac{x}{p}\right] \quad (27)$$

Where we have defined  $z = 1 + x$  and it understood that  $-p < x < p$  for the ingress/egress. For the purposes of the evaluating the maximum error, we know that  $x \simeq p$  and thus we may expand the cosine term into second order using a Taylor series. Let us assume we have the simple case of  $b = 0$  which means that  $i = \pi/2$ . We make further small-angle approximations to simplify the resultant expression for the error, which is justified since  $2\pi t_I \ll P$ . The other adjustment we need to account for is that so we have approximated  $b = 0$  and  $e = 0$ . To generalize the result, we consider that the effect of  $b > 0$  and  $e > 0$  is to stretch or shrink the ingress/egress duration by a factor  $\tau/\tau_0$ . Therefore our expressions here are actually for  $\mathcal{I}_0$ , which may be written as  $\mathcal{I}_0 = \mathcal{I}(\tau_0/\tau)$ . We may now rewrite equation (24) as:

$$\sigma_{\hat{F},\text{obs}} > \left| \frac{\psi^{5/2}}{108m^3} \left[ 3\left(\sqrt{24mp - 9\psi} - 4\sqrt{2mp - \psi} - 6\sqrt{4mp - \psi} + \sqrt{8mp - \psi}\right) \right] \right| \quad (28)$$

Where we have used:

$$\psi = \frac{2\pi a_R}{P} \frac{\tau_0}{\tau} \mathcal{I} \quad (29)$$

$$\frac{\tau_0}{\tau} \simeq \frac{\sqrt{1 - b^2} \sqrt{1 - e^2}}{q_c} \quad (30)$$

Due to the approximations made, we find that this equation is only stable for  $m \geq 2$ . For any given data set, we simply need to solve equation (28) for  $m$  with some sensible estimates of  $p$ ,  $b$ ,  $e$ ,  $\omega$ ,  $a_R$  and  $P$ . As an example, for Kepler-5b, taking the quoted parameters from the Koch et al. (2010) paper, we find that even using  $m = 2$  provides an error of 0.1ppm, which is well below the typical measurement uncertainty of 130ppm.

### 3.3.2 Limb-darkened trough-curvature

Another part of the lightcurve where we have significant curvature, and thus expect the maximum numerical integration errors, is the limb-darkened lightcurve trough. However, the peak-to-peak size of the changes in flux induced by the limb darkening are much lower than the transit signal itself (i.e.  $\delta$ ); typically an order-of-magnitude. Further, the time scale over which these changes act is greater than that of the ingress/egress curvatures (i.e.  $t_F \gg \tau$ ) except for grazing transits. So we can see that, in general, the errors in our numerical integration techniques will be dominated by the ingress/egress curvatures rather than the limb-darkening-induced lightcurve-trough curvatures.

### 3.3.3 Contact point discontinuities

The final source of variation in the lightcurve gradient is that of the discontinuous change located at the contact points. Estimating the error due to this discontinuity is most easily estimated by assuming a trapezoid approximated lightcurve and considering the location where maximal error is induced. The largest error (and in fact only error) will occur for measurements close to contact points, or more specifically  $|t_i - t_M| < \mathcal{I}/2$  where  $t_M$  is the time of one of the contact points.

Before the first contact point, we have a flat line at  $F = 1$  and after this point we have a linear slope with a gradient  $-(\delta/\tau)$ . The error in Simpson's composite rule will depend upon the relative phasing between the centre of the integration and the contact point, i.e.  $(t_i - t_I)$ . Generalized to any phase, the true integrated flux of the trapezoid approximated lightcurve for the  $i^{\text{th}}$  time stamp is given by:

$$\tilde{F}_{\text{true},i} = 1 - \frac{\delta}{\tau} \frac{(2t_i + \mathcal{I})^2}{8\mathcal{I}} \quad (31)$$

For each value of  $m = 1, 2, 3, \dots$  we choose to set the phase to be such that the difference between the true integrated flux and that from Simpson's method is maximized. Under such a condition, it may be shown that the maximum error is given by:

$$\sigma_{\tilde{F}}^{\text{Comp. Simp.}} = \frac{\delta}{\tau} \frac{\mathcal{I}}{24m^2} \quad (32)$$

$$= \frac{\delta}{\tau} \frac{\mathcal{I}}{6(N-1)^2} \quad (33)$$

For the system parameters of Kepler-5b, we find that using  $m = 1, 2, 3$  induces a maximal error of 371ppm, 93ppm and 41ppm respectively. Given that the measurement uncertainties are 130ppm (Koch et al. 2010), a suitable choice for the resolution would be  $m = 2$  since this means the maximum possible error of a data point in the least-favourable phasing would be below the measurement error.

It is interesting to see that for  $m = 2$  the error was 0.1ppm for the ingress/egress curvature of the same system, suggesting the discontinuity error dominates the error budget. Actually, this is expected from the arguments made earlier in this paper. Therefore, in most applications, a selection for  $m$  based on the error induced by the contact point discontinuities will provide a robust integration resolution.

### 3.4 Resampling

An additional method for numerically integrating the lightcurve is discussed here. Let us consider that we have observations with integrated time stamps given by the vector  $\tilde{t}$ . A second way of calculating  $\tilde{F}(\tilde{t})$  is to resample the time vector into a very fine cadence, at which point we may assume  $\tilde{F} = F$ . Let us define our temporary resampled time vector as  $\tilde{t}'$ . As an example, for the *Kepler* data, we may choose to resample the 30 minute integrations into 1 minute integrations would be done by expanding each time stamp,  $\tilde{t}_i$  into a sub-vector of 30 equally spaced time stamps with a mean value given by  $\tilde{t}_i$ . Our new temporary time array is used to generate a lightcurve using the normal MA02 expressions giving us  $\underline{F}'(\tilde{t}')$  (note that  $F$  here has no tilde because the MA02 equations can only generate instantaneous flux, not integrated flux). We then rebin the model lightcurve back to the original cadence to give  $\underline{F}(\tilde{t})$ . Finally, we make the assumption  $\tilde{F}(\tilde{t}) \simeq \underline{F}(\tilde{t})$ , i.e. the high cadence resampled time vector yields a lightcurve model consistent with a time vector of infinite cadence.

$$\tilde{t} \xrightarrow{\text{resample}} \tilde{t}' \xrightarrow{\text{MA02}} \underline{F}'(\tilde{t}') \xrightarrow{\text{rebin}} \underline{F}(\tilde{t}) \simeq \tilde{F}(\tilde{t}) \quad (34)$$

It can be seen that resampling into  $N$  sub-time stamps will increase the computation time by a factor of  $\sim N$ , since typically the MA02 subroutine uses the majority of a lightcurve fitting algorithm's resources (especially for non-linear limb darkening). In the next subsection, we will show that the computation times can be decreased by *selective resampling*.

One advantage of resampling is that we can choose to resample in such a way as to account for read-out and dead-times, which may be important if the instrument's duty cycle is quite poor<sup>1</sup>. The resampling of the  $i^{\text{th}}$  time stamp into  $N$  sub-time stamps with labels  $j = 1, 2, \dots, N-1, N$  can be expressed as:

$$t_{i,j} = t_i + \left(j - \frac{N+1}{2}\right) \frac{\mathcal{I}}{N} \quad (35)$$

The flux of the  $i^{\text{th}}$  time stamp is found by rebinning all  $N$  flux stamps from  $j = 1$  to  $j = N$ .

$$\tilde{F}_i = \frac{\sum_{j=1}^N F_{i,j}}{N} \quad (36)$$

Thus for the first few values of  $N = 2, N = 3$  and  $N = 4$  we would have:

$$\tilde{F}_i(N=2) = \frac{1}{2} [F(t_i - \mathcal{I}/4) + F(t_i + \mathcal{I}/4)] \quad (37)$$

$$\tilde{F}_i(N=3) = \frac{1}{3} [F(t_i - \mathcal{I}/3) + F(t_i) + F(t_i + \mathcal{I}/3)] \quad (38)$$

$$\begin{aligned} \tilde{F}_i(N=4) = \frac{1}{4} [ & F(t_i - 3\mathcal{I}/8) + F(t_i - \mathcal{I}/8) \\ & + F(t_i + \mathcal{I}/8) + F(t_i + 3\mathcal{I}/8) ] \end{aligned} \quad (39)$$

<sup>1</sup> We note that this is not the case for *Kepler* which has a duty cycle of 91.4%.

For a trapezoid approximated lightcurve, it can be easily shown that the error in these expressions, as a function of  $N$ , is given by:

$$\sigma_{\bar{F}}^{\text{Resampling}} = \frac{\delta}{\tau} \frac{\mathcal{I}}{8N^2} \quad (40)$$

Therefore, the resampling method yields greater accuracy than the composite Simpson's method. In Kipping & Bakos (2010), both the resampling and Simpson's composite rule were employed in completely independent analyses and the obtained results were consistent. Therefore, whilst we are free to use either method discussed here, the most efficient approach out of the two is resampling.

Gilliland et al. (2010) reported that they used a method for fitting the long-cadence lightcurve of TreS-2b which we interpret to be effectively equivalent to the resampling method. The authors split the LC intervals into 30 contributing sub-intervals corresponding to the SC cadence i.e.  $N = 30$ . For the reported LC r.m.s. noise of 66ppm and the system parameters of TreS-2b taken from Winn et al. (2008), we estimate that using  $N = 5$  would produce a maximum possible error in the most unfavourably phased data point of 59ppm and thus using  $N = 30$  is excessive for this lightcurve. Our equations therefore permit for a reduction in computational time of 600%.

### 3.5 Selective Resampling

Resampling time stamps which satisfy  $|t_i - t_C| > (t_T + \mathcal{I})/2$  and  $|t_i - s_C| > (s_T + \mathcal{I})/2$  is unnecessary since  $\bar{F}_i = F_i = 1$  in such cases (assuming we have folded multiple transits about the orbital period). Note that we define  $s_C$  as the mid-time of the secondary eclipse and  $s_T$  as the duration from contact point I to IV of the secondary eclipse. We label this method of optimization as *selective resampling*.

Since  $(t_T + s_T)/P \sim 2/(\pi a_R)$ , this can reduce the number of time stamps which require resampling by an order of magnitude for continuous staring telescopes like *Kepler* and *CoRoT*. It should be noted that selective resampling will not be possible if the lightcurve model includes phase variations of the planet e.g. HAT-P-7b, Borucki et al. (2009).

## 4 CONCLUSIONS

We have explored how long-cadence data, with particular focus on *Kepler*, causes severe systematic errors in the retrieved physical parameters, unless accounted for. The effect is valid for any finite exposure time but increases with longer cadences. Long-cadence data smears out the lightcurve morphology, which acts to stretch out the ingress/egress duration and suppress limb darkening in the lightcurve trough. These two effects act to increase and decrease the retrieved impact parameter respectively. Critically, overestimating the impact parameter is shown to lead to severe underestimations of the stellar density which could lead to planetary candidates being rejected on the basis of being unphysical.

Numerical integration techniques permit for improved modeling of the transit lightcurve. We discuss two particular methods, the composite Simpson's method and resampling.

We provide expressions for estimating the errors of these techniques and find that both methods produce an error which scales as  $N^{-2}$  where  $N$  is the numerical resolution of the techniques. Out of these two discussed methods, the resampling approach yields a greater efficiency.

## ACKNOWLEDGMENTS

DMK is supported by STFC, University College London, HATNet and the Harvard Smithsonian Center for Astrophysics. Special thanks to Gaspar Bakos for technical discussions on this subject. Thanks to G. Tinetti, J. P. Beaulieu, M. Swain, I. Waldmann and P. Deroo for useful comments in preparing this manuscript.

## REFERENCES

- Agol, E., Steffen, J., Sari R. & Clarkson W. 2005, MNRAS, 359, 567
- Borucki, W. J. et al., Science, 325, 709
- Borucki, W. J. et al., Science, 327, 977
- Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S. & Winn, J. N., 2008, ApJ, 689, 499
- Carter, J. A. & Winn, J. N., 2010, ApJ, 709, 1219
- Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJL, 529, L45
- Deming, D., Harrington, J., Seager, S. & Richardson, L. J., 2006, ApJ, 644, 560
- Gilliland, R. L. et al., 2010, ApJL, 713, 160
- Henry, G. W., Marcy, G. W., Butler, R. P. & Vogt, S. S., 2000, ApJ, 529, L41
- Holman, M. J. & Murray, N. W., 2005, Science, 307, 1288
- Kipping, D. M., 2008, MNRAS, 389, 1383, (K08)
- Kipping, D. M., 2009a, MNRAS, 392, 181
- Kipping, D. M., 2009b, MNRAS, 396, 1797
- Kipping, D. M., 2010, MNRAS, submitted
- Kipping, D. M. & Bakos, G. A., 2010, ApJ, submitted
- Koch, D. et al., 2010, ApJL, 713, 131
- Mandel, K. & Agol, E. 2002, ApJ, 580, L171, (MA02)
- McKeeman, W. M., 1962, Communications of the ACM, 5, 604
- Seager, S., & Mallén-Ornelas, G., 2003, ApJ, 585, 1038
- Seager, S., & Hui, L., 2002, ApJ, 572, 540
- Tinetti, G., Vidal-Madjar, A., Liang, M.-C. et al., 2007, Nature 448, 169
- Winn, J. N., et al. 2005, ApJ, 631, 1215
- Winn, J. N., et al. 2008, ApJ, 682, 1283
- Winn, J. N., 2010, *Transits and Occultations*, EXOPLAN-ETS, University of Arizona Press; ed: S. Seager

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.